Note for 01/07/2021

Taewan Kim

January 7, 2021

Contents

1 What to do **1**
1.1 Plan for this week 1

2 Copula **2**
2.1 Definition of Copula 2
2.2 Covariance Matrix 2

3 Code **2**

1 What to do

1.1 Plan for this week

1. Make three different distributions with distinct copula method, corresponding to three cell types.
2. Combine the three datasets as one, and plot a matrix of scatterplots
3. Check the shape of each plot if it resembles the cell data.
4. If not, try with different μ and Σ for each copula to make a L shape, linear, and etc.

2 Copula

2.1 Definition of Copula

Gaussian Copula is a method to transform the random variable $X = (X_1, ..., X_p)$ to a new random variable $f(X) = (f_1(X_1), ..., f_p(X_p))$ with assumption that $f(X)$ is multivariate Gaussian.

- A nonparametric extension of the normal, **nonparanormal**, depends on the functions f_j, and a mean μ and covariance matrix Σ
- When f_j are monotone and differentiable, $NPN(\mu, \Sigma, f)$ is a Gaussian Copula.
- If $X\ NPN(\mu, \Sigma, f)$ is nonparanormal and each f_j is differentiable, then $X_i \perp\!\!\!\perp X_j | X_{\setminus\{i,j\}}$ iff $\Omega_{ij} = 0$ where $\Omega = \Sigma^{-1}$

2.2 Covariance Matrix

Covariance matrix, $\Sigma_{jk} = Cov(X_j, X_k) = E[(X_j - E[X_j])(X_k - E[X_k])] = \sigma_{jk}$, is a square matrix

- The inverse of the covariance matrix is called "precision matrix", Ω.
- Σ is symmetric. (i.e., $Cov(X_i, X_j) = Cov(X_j, X_i)$)
- Σ is positive semi-definite (PSD).

$$u^T \Sigma u = \sum_{i,j=1}^{n} u_i \Sigma_{ij} u_j \quad = \sum_{i,j=1}^{n} Cov(u_i X_i, u_j X_j) = Cov(\sum_i u_i X_i, \sum_j u_j X_j) \geq 0$$

- Σ's eigenvalues are non-neagtive because Σ is PSD.

3 Code

- For the setting, the numbers of observations in first cell, second cell, and third cell are 150, 150, 200.
- The number of features are 8.

1. First try

   - First Cell Type
     Mean: 0, Cov: 0.1 but [(1, 3), (1, 4), (5, 6)]: 0.7, f: $1.2 * sign(x_i) * abs(x_i)^2$
   - Second Cell Type Mean: 0, Cov: 0.3 but [(1, 3), (1, 4)]: 0.7, [(5,7), (5,8)]: 0.6, [c(2,3), c(6,7)]: 0.4, f: $1.5 * sign(x_i) * abs(x_i)$.
   - Third Cell Type Mean: 0, Cov: 0.5 [(5,6)]: 0.2, [(4,7)]: 0.3, f: $0.8 * sign(x_i) * abs(x_i)^{0.8}$